DOCUMENT RESUME

ED 357 375 CS 213 830

AUTHOR McCurry, Niki
TITLE The Computerized Inventory of Developmental Writing Traits.
PUB DATE 28 Feb 92
NOTE 6p.; Several lines of very small print may not reproduce legibly.
PUB TYPE Reports - Descriptive (141)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Computer Assisted Testing; Elementary Secondary Education; Higher Education; Program Descriptions; Student Evaluation; *Test Construction; *Test Reliability; *Test Validity; *Writing Achievement; *Writing Evaluation
IDENTIFIERS Alaska Writing Program; Process Approach (Writing)

ABSTRACT

The Computerized Inventory of Developmental Writing Traits (CIDWT) is meant to provide a valid reliable measure of program improvement, particularly for teachers implementing a process writing approach in their classrooms. While standardized tests, portfolio, and holistic scoring all have something to offer, the CIDWT is an inexpensive direct performance assessment instrument which yields normed aggregatable data. The CIDWT, developed by a research team from the Alaska Writing Program, combines traditional numeric count research and computer based research; it is a MS-DOS computer program that counts and analyzes targeted numeric indicators in text files. Reliability and validity studies indicated that: the same four factors emerged in all studies (fluency, sentence development, word choice, and paragraph development); the CIDWT correlates well with teacher ratings; all factors and correlations were as strong or stronger for elementary students as for college students; and the CIDWT measures the development of writing traits. The next step in the research process is to collect a large national sample of student writing and combine that with the CIDWT results into a national database. (Contains 17 references.) (RS)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

February 28, 1992

The Computerized Inventory of Developmental Writing Traits

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Niki McCurry
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Niki McCurry, Ph.D.

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
✓ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.

ED357375

Are you implementing a writing process approach in your language arts class? Are you happy documenting the success of your writing program with standardized tests such as ITBS or CTBS? Do you believe that your students are learning to become better, more fluent, more organized, more expressive writers?

If you answered Yes, No, Yes, you should read this paper.

1. CIDWT Goals

Grading student compositions has always loomed large in the minds of American teachers. Back in 1859 Oliver Wendell Holmes Sr. (father of the famous jurist) drew upon the popular folk image of the hard-working teacher slaving away correcting student papers in his best selling potboiler, Elsie Venner. For 150 years that image has never left us. If students write papers, we mark 'em. MARKS 'R US. Why? Because in both the minds of teachers and the American public, that's what teachers do, they grade (read "correct" or "red mark" or "mark") papers. It's like breathing. Reflexive. Let's consider why teachers mark or grade or evaluate student writing.

First we mark papers to give grades. We have to give some kids A's, some B's, and some C's and so on. If we gave every kid an A, the world would probably come to an end, or at least the school system. Grades reward the good kids and punish the "bad." Ultimately grades allow schools to rank children, to decide who goes to Harvard, who to the army and who to the street. Ranking students is a bureaucratic function. Schools serve that function. But grades per se do not serve any instructional purpose.

When teachers accompany a letter grade with suggestions for improvement, then an instructional goal can be identified. The student can take those suggestions on paper #1 and use them hopefully to do a better job on paper #2. That doesn't happen often. More powerfully, if the student is allowed to use the teacher's suggestions on paper #1 to rewrite and improve paper #1b, the second draft, the suggestions can really help a student to improve. In the second instance, the teacher is serving as a coach, helping the student to improve his writing performance. Coaching is very powerful and both research and my own experience supports it. Coaching through written comments and one on one conferences really helps kids. It is a very justifiable reason for "grading" papers.

Finally, teachers grade or evaluate student writing, especially in groups, to find out how well the writing program is working. In this case, we are looking at the student outcomes as a measure of how good a job we are doing as teachers. If the goal of a writing program is to have kids write well, then one does need to examine how good the student product is. If the teacher can summarize or aggregate the evaluations of the writing of a whole classroom of young writers or the principal aggregate the evaluations of all the classrooms in her school, or the language arts supervisor all the schools in the district, then each of those professionals can get a handle on how well the program is going. If a new program has been implemented, or new training, or a new textbook or computer-based curriculum put in place , then such aggregated "grades" can measure program gains or losses, program strengths, and program weaknesses. This is the program improvement function of "grading papers."

So there are three very distinct reasons for grading papers: to rank students, to coach students, and to measure curriculum improvements. It is this last goal which CIDWT project addresses. The CIDWT does not seek to replace teachers either in their role as evaluators and graders of students or as coaches and helpers of students. The value of mere ranked grading of student writing is questionable unless it is accompanied by coaching. Teacher one-on-one coaching of students is of proven value (Hillocks). The Alaska Writing Program, the parent project of the CIDWT research, includes several coaching modules for that very reason. Both of these activities, ranking and coaching pose significant issues for discussion in the world of writing instruction. But when we talk about scoring student papers, people tend to muddle the three issues. Since the CIDWT is a computerized measure, we thought it important to specify which "grading" goal the project addresses. So relax your fears of computers "grading" kids or your hopes of getting out from under the burden of reading student essays. CIDWT does neither. It is meant to provide a valid reliable measure of program improvement, the program evaluation function.

2. Zeroing in on Program Evaluation

So we agree that we want to evaluate our writing program's effectiveness. We want to know what is working and what's not. What choices do teachers, principals or superintendents have to evaluate how well their students are learning to write? Up until now, schools have had three choices, each with its benefits and drawbacks. The CIDWT research offers a fourth or additional choice.

The easiest way to measure the writing program is the standardized test such as the *SRA*, *CTBS*, or *ITBS*. These measures have much to recommend them. They are cheap, costing only about $6.50 per students for almost 12 subtests. They are valid and reliable normed, easy to administer, and the public believes in them. It is easy to aggregate the results, to

CS213830

get data not just on one child, but to summarize results for a whole class, a whole school, a whole grade, a whole district. Simple numbers tell the public how good or how bad its schools perform in basic skills. But standardized tests have limitations. It is not at all clear that the "language expression" or "usage" subtests measure writing skills. These subtests probably measure children's ability to proofread for conventions such as punctuation or their familiarity with the preferred "school" dialect as opposed to their home dialect. But teachers throughout the land who have been working to develop their students' writing fluency, diction, organization, creativity, and expository skills are unsatisifed with the standardized tests as a measure of what is going on in their classrooms. The writing process approach being embraced by teachers in increasing numbers is probably not reflected in standardized tests. More conclusively, the tests sample student proofreading behaviors on other people's writing; they do not measure the student's own writing. Standardized tests are not *direct performance measures*; they do not evaluate the student actually performing the task to be evaluated. Tests measure correctness and social dialect and not much else. Yet good writing is not merely correct writing. There is so much more.

In response to the need for less simplistic data teachers developed the concept of *Portfolios*. Portfolios are great. Portfolios provide very useful very rich information for individual children, teachers, and parents. Together in an end of the year conference, the parent, child and teacher can examine the range of writing experiences the child had, compare fall and spring writing samples, and review reading and literature as well. The portfolio helps the child gain metacognitive knowledge about how he/she learns or thinks. The child gets to select one best or most representative paper for evaluation. Portfolios help in instruction, parent-teacher, and teacher-child communiction. But by their very richness, portfolios do not really allow for aggregatable data. They are good for individuals but not for groups. How do you qualify, summarize, and average portfolio information for a whole class or school? How do you generalize and quantify portfolio information to make programmatic judgments?

Holistic scoring does yield quantifiable data. In holistic scoring, a group of teachers goes through a process of identifying benchmark papers and developing a scoring rubic so that they can in teams of two or three quickly give each paper an overall score from 1-5 or 1-6 or 1-10. That gives a number which has validity and reliability within the group. Using this method teachers can grade hundreds of papers in a day. Analytic trait scoring works the same way but focusses in on specific writing traits like style, creativity, and organization. Holistic scoring works very well. It has great value in training teachers to be clear about instructional goals and in articulating those goals to themselves and their students.

Even as a one time staff development exercise, participating in a large group holistic scoring activity has a great merit. As an instrument to measure program improvement, holistic scoring has great face validity: it measures real writing skills of real students engaged in really performing the task you want to test. It is a *direct performance measure*. But this procedure has its limitations. First, it is expensive, it costs about $5 per student and takes quite a bit of teacher time. Second, the scores, while quantifiable, are not normed. You can't compare one district's average score of 4.5 to another's of 6.1. You can't compare across districts, states, or schools. You can get a state level score if the writing assessment is conducted at the state level as occured in Oregon and California in 1991. But national scores and comparisons are impossible right now.

Each of these measures has something to offer. Standardized tests are simple, cheap and provide good numbers for crunching. Portfolios are rich sources of information and learning. Holistic scoring is a great teacher training measure. Most schools don't choose between the three. Many schools use all three measures, with everything ending up in the portfolio. But even in this case, something is still missing. *What's missing is an inexpensive, valid, reliable, direct performance assessment instrument which yields normed aggregatable data*. The CIDWT, *The Computerized Inventory of Writing Traits*, supplies the missing piece. CIDWT is a valid, reliable, normed instrument which uses the power of the computer to inexpensively assess student writing for purposes of program development.

CIDWT was developed by a research team from the *Alaska Writing Program*. The *Alaska Writing Program* is a nationally validated exemplary computer-based writing program disseminated through a federal grant to more than 35 school districts in six states. The evaluation team of Alaska Writing Program was faced with the task of providing to the funding agency, Title VII, aggregated valid reliable data on the impact of the Alaska Program on the writing of children in the 201 classrooms implementing the program. For the reasons discussed above, they reviewed and rejected all three alternatives. They needed an inexpensive, valid, reliable normed instrument which was a direct assessment measure. There wasn't any. So they invented it: *The Computerized Inventory of Writing Traits* .

3. CIDWT : Theoretical Backgrounds

CIDWT sounds like a very new creature in the testing jungle. But it has an old and honorable family history. Its two very unlikely parents are the NCTE and IBM. The CIDWT research team combined two ideas which have been available in the research since 1959 to come up with a revolutionary new concept and instrument: traditional numeric count research and computer based research.

Traditional numeric count research has been around since World War II. There is a strong tradition in the language arts field supporting research based on numeric indicators. The most widely known and accepted use of countable numeric indicators to stand for language use or comprehensibility or quality are the readability formulas, such as those of

Edward Frye and Rudolph Flesch. Both Flesch and Frye formulas are based on sentence length and word length. Walter Loban did a now famous study 13 year longitudinal study of children's language development in which he featured several counts to document growth of student language skills. In 1962, Walker Gibson, then president of NCTE, published his now classic rhetorical analysis of American prose style (focus on Fitzgerald, Hemingway, and Faulkner) using count formulas he included in the back of his book. Crucial to the development of the technique of sentence combining was the sentence complexity research of Kellogg Hunt, John C. Mellon and Frank O'Hare all of which depended on counting sentence length, occurrences of subordinations and other clause types, and most importantly, t-unit length. Richard Lanham's formality index and Lester Golub's sentence density research also depend on counts. All of the above research was incorporated into CIDWT . Ironically, although the National Council of Teachers of English has been very resistant to even thinking about the notion of computer-assisted writing assessment, much of the research behind CIDWT referenced above is NCTE-sponsored, supported, or published.

Parallel to but not dependent on the traditional numeric . research pursued by NCTE were early p neering efforts in the 60s using computers to assess writing. Ellis Page, one-time president of the American Educational Research Association, then at MIT, conducted the first research and articulated the theory of "Trins" and "Proxes." *Trins* are intrinsic qualities of writing such as organization, diction, and creativity; *proxes* are numeric indicators for these qualities. Thus, he theorized that the number of sentences using subordinations might be a *prox* for the *trin* of sentence complexity. He assumed that for each *trin* there was one *prox*. Page used statistical analysis to show the relationships between his *trins* and *proxes* and found statistically significant correlations between his indicators and human raters.

Patrick Finn replicated Page's research. Henry Slotnick articulated Page and Finn's work into a theoretical framework and carried the notion a step further, using more advanced factoral analysis techniques. He too found statistical signficance. More importantly, he found that factoral analysis revealed multiple numeric indicators for each trin. He suggested that one might also use the terms independent variable for proxes and factors for trins. Interestingly, Slotnick's statistical analysis revealed six factors: fluency, misspelling, diction, sentence structure, punctuation, and word choice.

This very successful computer-based research hit two walls. Humanists such as Ken Macrorie were horrified and repulsed by the notion of computers grading students. Moreover, these early researchers had no way of getting student writing into a machine readable form except for cumbersome key punching. In the sixties computers-assisted composition scoring had nowhere to go. So there the research sat until the Alaska Writing Program team found it hiding in the dusty ERIC descriptors. The AWP group drew upon this wealth of computer based and NCTE sponsored research to design the CIDWT .

4. The CIDWT Design & Function

The CIDWT functions very simply. It is a MS-DOS computer program which counts and analyzes targeted numeric indicators in text files. Student essays are word processing files fed to CIDWT in class batches of 30 or fewer. The pilot form, CIDWT 1.0b, simply counts the independent variables listed above and prints out raw score counts for each variable (also saves raw scores to disk as data files). CIDWT 2.0 (projected design, completion January 1993) will convert the raw scores to weighted scores, t-scores and norms. CIDWT 1.0b counts 35 independent variables which were selected based on the research discussed above. CIDWT 2.0 will count additional variables, such as breaking down the subordination count to each subordinating conjunction, and breaking down punctuation into seperate counts for each specific punctuation mark.

Variables

total words	total paragraphs¶
SD sentencelength	# punctuation
Av. word length	SD. word length
% unique wds	FOGG
Flesch Av. sent length	Av.¶ length
# prepositions	#To Be verbs
# articles	# coordinates
#subordinates	# conditionals
# opinion words	#vague words
#transitions	#pronouns
#slang words	#-ion words
# THEs	#most common words
%most common	#very common
%very common	#common
%common#semi-common	%semi-common
#uncommon	%uncommon

CIDWT runs on IBM or IBM compatible computers. Using a hard drive and a 486 microprocessor, CIDWT can generate raw scores for about 40-44 essays per minute. That is fast. Fast scoring means that ultimately CIDWT can provide very reliable data very inexpensively. Given the widespread use of word processors in the writing classroom today, it should be relatively simple and inexpensive for districts to collect student writing samples as word processing files. So long as the word processing file can be saved as a basic text file, regardless of what computer the original story was written on, it should be possible to transfer the file to MS-DOS for analysis by CIDWT. Technically, CIDWT works like a charm. If it can be demonstrated to measure real writing traits, CIDWT could be of great value to writing teachers as a program evaluation instrument. So the next question is: What does the CIDWT measure? Does it work? The CIDWT research team set out to answer those questions.

5. Reliability & Validity Studies

1989 Alaska Statewide 10th Grade (500 samples)
1990 CCNY College Freshmen 82 cases
1990 El Paso Community College 243 samples (Hispanic bilingual)
1990 San Jose State College Sophomores 75 samples
1990 Anchorage School District Grades 3, 6 & 8 (300 samples)
1991 Anchorage Study 904 samples

So we took the CIDWT for a test run around the mountain to see what we could see. Would it count? What would it count? Would we replicate the earlier research? Would it correlate to the ratings of real human being teachers?

Would it work for young writers or only for college students as had the earlier research? Are the Xs CIDWT measures developmental? Could it be normed? The list of reliability and validity studies demonstrates how very powerful CIDWT proved to be.

First, we did find great factoral consistency. *In every separate study, the same four factors emerged.* Note that with the exception of spelling and punctuation, thse are the same factors that Slotnick identified. More exciting, the same factors emerged time and time again in the same order. Factors are identified by the statistical process in order of power or strength. So not only were the factors the same in all six CIDWT studies, and Slotnick's earlier research, but the same factors were generated in the same order of power. The various samples used in the pilot studies had great variability in terms of where, who, how they were collected. Students ranged in age from 10 to 25, from Caucausian mainstream to Hispanic, Black, and Asian. Circumstances ranged from a casual collection of essays to a rigidily controlled timed writing sample. Samples ranged from written on computer to written by young children by hand, from stories to reports to descriptive essays. Regardless of circumstance, the same four factors emerged. We have called these factors writing traits and labeled them: fluency, sentence development, word choice, and paragraph development.

Factor I= fluency
word length
% unique words
Flesch
#common words***

Factor II=sentence development
S.D. sentence length
FOGG Readability
Av. sentence length***
Av. ¶ paragraph length
subordinates # conditionals
#-ion (nominalizations)
#the's

Factor III=diction/vocabulary development
Av. word length***
S.D. word length
%unique words
Flesch readability
#-ion words
#the's

Factor IV= Paragraph Development
Total ¶ # Flesch
Av. ¶ length #punctuations
#conditionals #opinion-words
-ion words SDsentence length
the's

It should be noted that the individual variables associated with each factor do vary depending on the age and size of the samples. As the sample size increases, the number of variables associated with that factor increases and seems to become richer and more compelling. But no experienced teacher reviewing the results of analysis of any of the samples would question the basic trait name for of each factor.

The *second* key result is that the CIDWT correlates very well and very consistently with teacher ratings of the same paper. That means that it is likely that a paper receiving a high CIDWT rating will receive a high score from a teacher. We found that the more consistent the human raters, the better correlation with CIDWT . We have R values as high as .81 for the third grade sample and .95 for the San Jose College sample. In every case tested we have found statistically significant correlation between CIDWT and teacher ratings.

A *third* concern was the applicability of the CIDWT to the writing of elementary school students. The initial computer-scoring research was all done at the college level. There was a real concern by the team that the variables measured would not show up in the writing of younger students. However, we found that the factors and teacher correlations were just as strong or stronger for younger students as for older students. CIDWT has a functional range from grade 3 to college sophomore.

Fourth, we found that CIDWT measures the development of writing traits. The factors develop at incremental rates from grade 3- 12. This means that CIDWT will allow us to trace the growth and development of a student or group of students' writing skills over their years in school.

1991 Pacific Coast Norms Study

Having conducted pilot studies to demonstrate that the CIDWT Inventory could measure the development of four key writing traits (fluency, sentence development, word choice, and paragraph development), the next step was to create norms for the scores. Raw scores tell teachers and children little. How good is a score of 403 on fluency? Norms tables convert the raw scores to percentile ranks within a meaningful range. Due to the n of the norms samples, it was not possible to develop grade level norms at this stage. Statistical surety demands more than 200 samples per level. By combining grades 3-5, 6-8 and 9-12, we were able to get enough samples in three levels to develop reliable norms for elementary school, middle school, and high school. That means that given the raw score from CIDWT the research can use the norms table to tell whether that paper is at the 75th percentile for an elementary school student, or 45th for high school and so on. So a given paper can now be compared on the four writing traits to its peer group within a three or four grade span.

6. National Data Base of Developmental Writing Traits Project

The next step in the research process is to collect a large national sample of student writing and combine those CIDWT results into a national data base. As more student writing samples are merged into this data base of writing trait scores, more information about the strength of the factors will emerge. For example, with a large enough data base, we can get information about how and when students begin to use certain sentence development skills like subordination. Eventually stronger norms tables can be developed which have a national basis.

The national data base will work this way. Districts which wish to participate will contact the project director. All student papers must be submitted on 3.5 inch disk on ASCII files, with student information form and hard copy of the paper. Participating districts will be sent bubble dot student information forms (similar to those used by standardized test makers) and a proofreading disk. The student information forms will be filled out by either teacher or student using no. 2 pencil to collect specific information about that student and that piece of writing, such as the student's name, grade, categorical status (bilingual, GT, Chapter 1), how many days spent on this paper, whether originally written on computer or pen and so on. The proofreader disk will scan and correct only for paragraphing and sentence end spacing. The student forms, a hard copy of the papers, and student essays on disk will be sent to the data base center in Carmel, California. Districts will receive normed scores on CIDWT in return for contributing to the national data base.

If you are interested in learning more about the National Data Base of Developmental Writing Traits Project, contact:

Niki McCurry
Alaska Writing Program
Box 309
Nenana, Alaska 99760
(1-800-348-1335)

7. Bibliography

Gibson, Walker. *Tough, Sweet, and Stuffy.* Bloomington, IN: Indiana University Press, 1966.

Golub, Lester. "Syntactic density score with some aids for tabulation." ED091 741, 1973

Finn, Patrick J. "Computer-aided Description of Mature Word Choices in Writing, " in *Evaluating Writing: Describing, Measuring, Judging* (Cooper, Odell, eds.) Urbana, IL: NCTE (1977). pp.69-89.

Hunt, Kellogg. *Syntactic Maturity in Schoolchildren and Adults.* Monographs of the Society for Research in Child Development, 1970, Vol 35, no. 3.

Lanham, Richard. *Revising Prose.* Charles Scribner's Sons, 1979.

Loban, Walter. *The Language of Elementary School Children.* NCTE Research Report No.1, 1963.

McCurry, Alan and Niki. "Writing Assessment for the Twenty-first Century," *The Computing Teacher* (April 1992) pp.35-37..

Macrorie, Ken. "Roundtable Review," *Research in the Teaching of English*, 1969, pp.228-236.

Mellon, John C. *Transformational Sentence Combining: a method for Enhancing the Development of Syntactic fluency in English Composition.* NCTE Research Report No. 10, 1969.

Nivette, James, McCurry, N. and others. *The CIWE 1991 Norms Study.* Unpublished monograph, Yukon Koyukuk School District, February, 1992.

O'Donnell, R. C. and others. *Syntax of Kindergarten and Elementary School Children: a transformational analysis.* NCTE Research Report, no. 8, 1967.

O'Hare, Frank. *Sentence Combining.* NCTE Research Bulletin No. 15, 1971.

Page, Ellis, "The Imminence of Grading Essays by Computer, " *Phi Delta Kappan*, January, 1966, 238-243.

Page, Ellis, "The Use of the Computer in Analyzing Student Essays," *International Review of Education*," XIV, No. 2, 1968, 210-225.

"Project Shows Computers can Grade Essays." *Education Computer News*, November 27, 1990, 3-4.

Slotnick H. B. "Toward a Theory of Computer Essay Grading, " *Journal of Educational Measurement*, IX, (1972): 253-63.

Wresch, William. "Computer Analysis of Student Essays—25 Years of Research," unpublished paper presented at the National Testing Network in Writing Conference, Montreal, March, 1990.

CIDWT Research Team

Statistical Design: James Nivette, Ph. D.
Writing Theory: Niki McCurry, Ph. D.
Programming: William Wresch, Ph. D.
Instructional Plan: Alan McCurry, M.L.S.